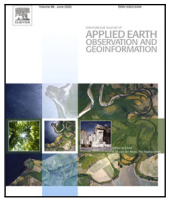
International Journal of Applied Earth Observation and Geoinformation 119 (2023) 103304

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Semi-supervised water tank detection to support vector control of emerging infectious diseases transmitted by *Aedes Aegypti*

Steffen Knoblauch [a,b,*], Hao Li [a,c], Sven Lautenbach [a,d], Yara Elshiaty [a,e], Antônio A. de A. Rocha [f], Bernd Resch [g,h], Dorian Arifi [g], Thomas Jänisch [i,j,k], Ivonne Morales [i,j], Alexander Zipf [a,b,d]

[a] GIScience Research Group, Heidelberg University, Heidelberg, Germany
[b] Interdisciplinary Centre of Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany
[c] Professorship of Big Geospatial Data Management, Technical University of Munich (TUM), Munich, Germany
[d] HeiGIT at Heidelberg University, Heidelberg, Germany
[e] Faculty for Mathematics and Computer Science, Heidelberg University, Heidelberg, Germany
[f] Institute of Computing, Fluminense Federal University (UFF), Niterói, Brazil
[g] Department of Geoinformatics, University of Salzburg, Salzburg, Austria
[h] Center for Geographic Analysis, Harvard University, Cambridge, USA
[i] Heidelberg Institute of Global Health (HIGH), University Hospital Heidelberg, Heidelberg, Germany
[j] Department for Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany
[k] Center for Global Health and Department of Epidemiology, Colorado School of Public Health, Aurora, USA

ARTICLE INFO

Dataset link: https://doi.org/10.11588/data/7LLXFP

Keywords:
Aedes Aegypti
Eco-epidemiology
GeoAI
Object detection
Ovitrap
Rio de Janeiro
Semi-supervised self-training
Urban health
Vector control
Water tank

ABSTRACT

The disease transmitting mosquito *Aedes Aegypti* is an increasing global threat. It breeds in small artificial containers such as rainwater tanks and can be characterized by a short flight range. The resulting high spatial variability of abundance is challenging to model. Therefore, we tested an approach to map water tank density as a spatial proxy for urban *Aedes Aegypti* habitat suitability. Water tank density mapping was performed by a semi-supervised self-training approach based on open accessible satellite imagery for the city of Rio de Janeiro. We ran a negative binomial generalized linear regression model to evaluate the statistical significance of water tank density for modeling inner-urban *Aedes Aegypti* distribution measured by an entomological surveillance system between January 2019 and December 2021. Our proposed semi-supervised model outperformed a supervised model for water tank detection with respect to the F1-score by 22%. Water tank density was a significant predictor for the mean eggs per trap rate of *Aedes Aegypti*. This shows the potential of the proposed indicator to enrich urban entomological surveillance systems to plan more targeted vector control interventions, presumably leading to less infectious rates of dengue, Zika, and chikungunya in the future.

1. Introduction

The recurring worldwide outbreaks of the severe acute respiratory syndrome (SARS)-associated coronavirus in 2003 and 2020 demonstrate how frequently and rapidly infectious diseases can spread in a globalized world. However, it is not globalization alone that is a driving factor for increased occurrences of infectious diseases but also climate change (Semenza et al., 2022). This is particularly true for mosquito-borne diseases, as rising global temperatures lengthens annual transmission seasons and leads to larger suitability areas for mosquitoes (Colón-González et al., 2021; Rocklöv and Dubrow, 2020). Considering all pathogen transmitting mosquitoes worldwide, *Aedes Aegypti* is the most prevalent one (European Centre for Disease Prevention and Control, 2016; Wilke et al., 2020). It is the primary vector for Zika, chikungunya, yellow fever, and dengue with a 30-fold increase in incidences over the last 50 years (Ebi and Nealon, 2016). The WHO is estimating that by 2080 over 60 percent of the world's population will live under direct risk of *Aedes Aegypti* (WHO, 2017; Messina et al., 2019). This turns this disease vector into an emerging global threat (Ebi and Nealon, 2016).

As of now there is no effective vaccine for dengue (WHO, 2022; Amorim and Birbrair, 2022; Schwartz et al., 2015), Zika, or chikungunya (Kantor, 2018; Schrauf et al., 2020). Accordingly, vector control, involving the process of eliminating vector breeding habitats and the application of insecticides to maintain mosquito populations at acceptable level, remains the most effective countermeasure for these diseases (Wilson et al., 2020; Hladish et al., 2020; Lobo et al., 2018).

* Corresponding author at: GIScience Research Group, Heidelberg University, Heidelberg, Germany.
E-mail address: steffen.knoblauch@uni-heidelberg.de (S. Knoblauch).

https://doi.org/10.1016/j.jag.2023.103304
Received 9 December 2022; Received in revised form 31 March 2023; Accepted 8 April 2023
Available online 19 April 2023

Vector control, however, is very costly as it requires a massive workforce and is often limited by regulative constraints on the use of insecticides, especially in urban areas where most infections by *Aedes Aegypti* occur (Knerer et al., 2020; Taborda et al., 2022). Therefore, mapping of *Aedes Aegypti* on the urban scale is of particular interest in order to implement local vector control measures in a more targeted manner and, above all, at lower costs (Runge-Ranzinger et al., 2014; Boser et al., 2021; Limkittikul et al., 2014; Da Queiroz and Medronho, 2022). This is especially important for the Global South where public health budgets for disease prevention are often scarce despite financial support from the WHO (Yukich et al., 2008).

The mapping of *Aedes Aegypti*, however, is not trivial. *Aedes Aegypti* is an urban favoring mosquito with a short flight range of around 200 meters and thus limited habitat size (Honório et al., 2003; Bomfim et al., 2020; Harrington et al., 2005). It lives in close vicinity to its breeding sites. These can be characterized as small artificial water containers, such as discarded tires (Getachew et al., 2015), buckets, barrels, pet dishes, construction blocks (Morrison et al., 2004; WHO, 2012), storm drains (Paploski et al., 2016), trash (Banerjee et al., 2015), flower pots (Vezzani, 2007), or water tanks (Trewin et al., 2021). Many of these containers occur with great spatial variance due to social urban structures (David et al., 2009). This, in combination with the small size and the limited flight range of mosquitoes, leads to a high spatial variability of *Aedes Aegypti* abundance. It differentiates *Aedes Aegypti* strongly from other mosquito species such as the malaria transmitting *Anopheles* mosquito, which tends to breed in large natural water bodies and thus occurs in higher concentrations (Chavasse, 2002; Gwitira et al., 2018; Youssefi et al., 2022). Consequently, the task of spatial mapping of mosquito distribution is more challenging for *Aedes Aegypti* than for other mosquito species (Boser et al., 2021).

Currently, there are two *Aedes Aegypti* mapping approaches in use to implement vector control in a more efficient and cost-saving manner. One of them are sample-based entomological surveillance systems including the positioning of mosquito traps and the conduction of household surveys (Pan American Health Organization, 2019; Bowman et al., 2014). These monitoring systems require a large amount of manual work but provide valid information on mosquito abundance, such as precise counts of mosquito eggs, larvae, and pupa. Nevertheless, they are hard to scale due to their labor-exhaustive nature, often cannot cover larger areas in high resolution and need trained personnel, which all together limits its practical scope (Vasconcelos et al., 2021). The alternative approach is based on the modeling of mosquito abundance via spatial proxies using modern computing techniques and harnessing big spatial data such as satellite imagery. These methods are less precise, because proxies by definition only provide indirect evidence of a phenomena. However, they require less manual work and consequently offer the possibility for a much broader spatial coverage as well as continuous mapping to capture the high spatial variability of *Aedes Aegypti* abundance in urban areas (Boser et al., 2021; Louis et al., 2014).

There are several approaches in literature showing the benefits of *Aedes Aegypti* mapping with spatial proxies. Some studies retrieve proxies from citizen science (Caputo et al., 2020; Cho et al., 2021; Low et al., 2021, 2022; Muñoz et al., 2020; Agarwal et al., 2014), others from remote sensing (Lorenz et al., 2020; Cunha et al., 2021; Dandabathula, 2019; Fernandes et al., 2020; Chang et al., 2009; McFeeters, 2013; Machault et al., 2014; Uusitalo et al., 2019), street view (Andersson et al., 2018; Haddawy et al., 2019; Su Yin et al., 2021), or drone imagery (Passos et al., 2020; Dias et al., 2018; Haas-Stapleton et al., 2019; Mehra et al., 2016; Schenkel et al., 2020). However, these approaches have shortcomings. Remote sensing studies, based on not very-high-resolution (VHR) satellite imagery, are presumably too coarse to detect small scale features that provide breeding habitats. Approaches that rely on citizen science, drone imagery, or street view are hard to transfer to other case studies since the data used is only available for selected sites or expensive to get. Put differently, most of these approaches neither derive high resolution proxies with open accessible data, nor use scalable workflows to map *Aedes Aegypti* below 200 m to consider limited mosquito flight ranges (Louis et al., 2014; Sallam et al., 2017). The application of object detection models for urban *Aedes Aegypti* breeding site mapping based on open accessible satellite imagery is rare. The same applies for the combination of entomological surveillance data with automatic mapping workflows in the field, although there is a WHO pillar of action called "scale up and integrate tools and approaches for global vector control" (WHO, 2017). This paper addresses the following research gaps:

- **Research Gap 1:** There is a **need for high resolution proxies** to enrich entomological surveillance data of *Aedes Aegypti* to conduct vector control with more focus and lower costs.
- **Research Gap 2:** There is a **need for scalable approaches based on open accessible data** to retrieve high resolution proxies for the mapping of inner-urban *Aedes Aegypti* distributions.

We analyze how far recent advances in deep learning can help to close these research gaps. We envision that these can be applied to create both, more scalable and also more precise methods for the mapping of urban *Aedes Aegypti* abundance to support vector control. From our perspective advances that allow to capture small *Aedes Aegypti* breeding containers with low manual labeling effort are particularly promising. Especially the semi-supervised self-training (SSST) of object detection networks, as SSST can reach similar prediction performance as supervised methods but with less manual labels (Rosenberg et al., 2005).

In this paper, we demonstrate the potential of SSST for the mapping of urban *Aedes Aegypti* distributions. We apply deep learning based object detection models on VHR satellite imagery and design a SSST-based fine-tuning algorithm to map the density of rainwater tanks, a typical *Aedes Aegypti* breeding spot in urban areas of the Global South. We evaluate water tank density as a high-resolution proxy to model entomological surveillance data originating from mosquito traps. For our approach, we exclusively use open accessible data. This increases the applicability of our workflow for real-world scenarios. More specifically, we address the two following research questions:

- **RQ1:** To what extent can semi-supervised self-training outperform supervised learning methods with equal labeling effort for water tank detection?
- **RQ2:** How well does water tank density capture the observed inner-urban distribution of *Aedes Aegypti* in the case study?

2. Experimental design

In order to answer the derived research questions, we propose a novel framework for the semi-automatic mapping of water tanks (cf. Fig. 1). Our concept consists of mainly three parts: first open accessible input data to increase the transferability of our experiment, second the supervised and semi-supervised self-training of water tank detection models to evaluate the usefulness of SSST over supervised learning to support labor-intensive public health practices like entomological surveillance, and third large-scale water tank predictions to evaluate the significance of water tank density as a high resolution proxy to model inner-urban *Aedes Aegypti* distributions presumably useful for a more targeted planning of vector control interventions in the future.

2.1. Materials

2.1.1. Study site

We applied the proposed workflow to the city of Rio de Janeiro, which is one of the highest effected mega cities for mosquito borne diseases worldwide (Gibson et al., 2014; Wilson, 2011). The city belongs to the endemic regions for *Aedes Aegypti* transmitted diseases due to its year long tropical climate (Franco dos Santos et al., 2022). With a population of around 6.75 million people and a high connectivity

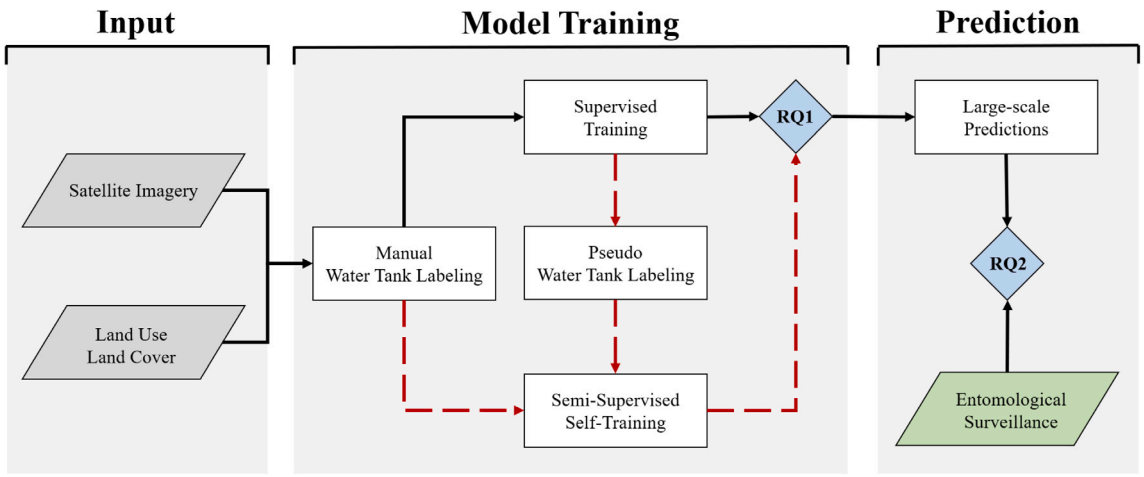


Fig. 1. Overview of the proposed framework to support future vector control including the required open accessible study data (gray), the semi-supervised self-training loop for water tank detection model fine-tuning (red), the evaluation of stated research questions (blue), and the ground truth evaluation set (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to other urban areas in Latin America, the second biggest city of Brazil has often been a starting point for larger uncontrolled disease outbreaks (Figueiredo, 2004; Da Silva et al., 2002). The proximity of different types of urban structure, such as favela and other residential areas, and the topography of the city account for a high variability of possible *Aedes Aegypti* breeding sites. This makes the city of Rio de Janeiro an interesting use case for our proposed method.

2.1.2. Study object

Our study object are water tanks, often used for drinking water storage in the city of Rio de Janeiro and other Latin American cities. Water tanks are known to be one of the main breeding spots for *Aedes Aegypti* (Trewin et al., 2021). In the city of Rio de Janeiro they are part of vector control measures as well as entomological surveillance systems (Secretaria de Vigilância em Saúde, 2013). However, by far the majority of water tanks are not monitored due to the labor intense process. Since the locations of water tanks in the city of Rio de Janeiro are not mapped, it was so far not possible to investigate the relationship between water tank presence and mosquito abundance. Water tanks usually have a radius of 1 to 2 meters and are objects in the format of a cylinder with a approximated height of 1.5 m. They appear dark blue, whereas some older ones might appear with brighter color due to sun bleaching (cf. Fig. 2).

The urban appearance of water tanks can be strongly correlated with the social structure of cities as shown by Cunha et al. (2021). As expected and further revealed by visual inspection, this was also the case for the city of Rio de Janeiro, where water tanks appeared more frequently in socially weaker parts of the city such as favelas due to the lack of piped water access. Thereby, the close proximity of favelas and other urban structures in the city of Rio de Janeiro leads presumably to a high spatial variance of water tanks. When observed, water tanks were primarily located in backyards and on rooftops. Due to their complex installation and heavy weight, it can be assumed that the position of water tanks does not vary much over time. All this makes the object detection of water tanks based on satellite imagery an interesting task to close the targeted research gaps.

2.1.3. Study data

We used three datasets for our approach, namely: satellite imagery, land use land cover (LULC) maps, and entomological surveillance data (cf. Fig. 1). Satellite imagery downloaded from the Microsoft Bing Tile Map Service (TMS) API (Microsoft, 2022) served as our main input data source. For the high resolution detection of water tanks we chose the highest available zoom level of 22 with an image resolution of 0.0373 meter per pixel. We retrieved 10,668,699 image patches of 256 × 256 pixels. In addition, we used administrative LULC data (Municipality of Rio de Janeiro, 2022) to derive information about the location and size of different urban structures, which was used to stratify water tank labeling (cf. Fig. 6).

As evaluation data we had access to an entomological surveillance database from January 2019 until December 2021 provided by the health ministry of Rio de Janeiro. The purpose of entomological surveillance is to monitor the distribution and impact of vector control measures. The reliability of this data is highly affected by the spatial coverage and temporal execution frequency. Examples for entomological surveillance measurements are mosquito count or index data for various development stages of *Aedes Aegypti*: eggs, larvae, pupa, and adult. For our use case we used data collected with 1,207 ovitraps distributed around the city (cf. Fig. 3). These are traps filled with water of around 20 cm radius used in the city of Rio de Janeiro for the collection of *Aedes Aegypti* eggs and larvae. When mature, the mosquitoes cannot escape these traps. The amount of eggs and larvae was collected on a monthly basis. As an evaluation indicator for our proposed method we used the "mean eggs per trap" (MET) rate aggregated over monthly time steps for *Aedes Aegypti*.

2.2. Methods

2.2.1. Manual water tank labeling

Since water tank locations were not available in sufficient amount and quality from open datasets, manual labeling was necessary. We manually labeled 2,000 water tanks in favelas and another 2,000 in residential areas for the training purpose. Labeling was done in QGIS (QGIS Development Team, 2022). The strata (favela/residential areas) were derived from the LULC map (Municipality of Rio de Janeiro, 2022). We labeled another 1,000 water tanks for validation of the model. Unlike the training labels, the labels for the test data were sampled across all kind of urban structures to analyze the robustness and generalization of our object detection model. All manual labels including are provided in the supplementary material.

2.2.2. Supervised training

We trained a Single Shot Multibox Detection (SSD) Network (Liu et al., 2015) on the manual water tank labels. The SSD was retrieved from the TensorFlow Object Detection Model Zoo API (Google, 2022). SSD networks are single-stage object detector architectures that have been successfully applied for the detection and mapping of geospatial objects of diverse size and shape. In addition, they offer a good balance

Fig. 2. Rainwater tanks for water supply in the city of Rio de Janeiro occurring in high spatial variability due to different urban structure types like residential areas (bottom left) and favelas (bottom right).

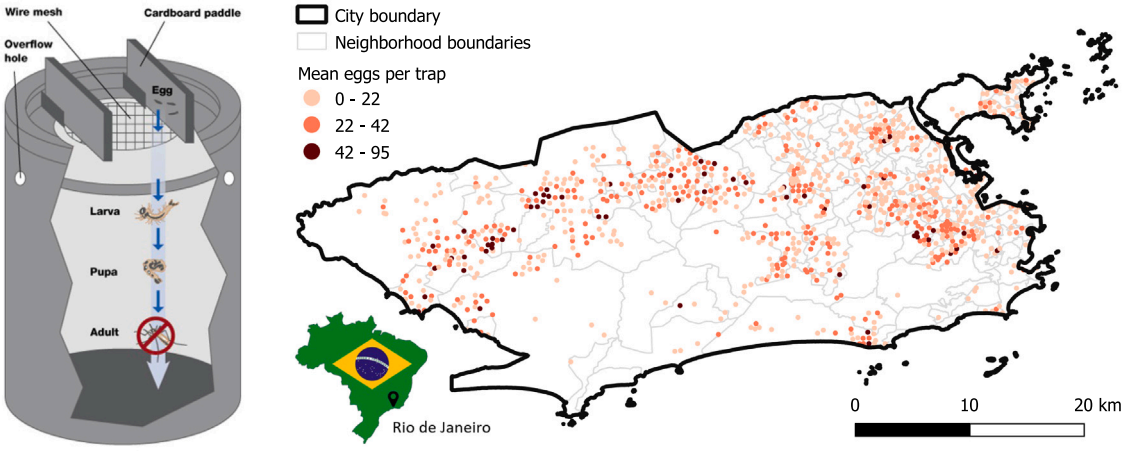


Fig. 3. Sketch of the ovitrap used by the entomological surveillance system (left) and corresponding locations of the traps in the city of Rio de Janeiro. The color indicates the "mean eggs per trap" (MET) rate for *Aedes Aegypti* between January 2019 and December 2021 (right.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between training time and accuracy when compared to two-stage object detection networks like Faster R-CNNs as shown in Model Zoo (Google, 2022). The output of an SSD network is a list of predicted features and the corresponding probability scores.

The SSD network we used consists of mainly two parts (see Fig. 4). First a feature extractor, which is in our case a backbone network with 164 layers, namely the Inception-ResNet-V2 network, allowing the model to learn deterministic features (e.g. colors and shapes) for common object detection tasks (Liu et al., 2015). Second a multi-layer detector together with a Non-Maximum Suppression (NMS) layer was used to create multi-scale detection boxes and to calculate the confidence scores of classification. This was needed for calculating the training loss. We used a SSD network pre-trained on the Microsoft COCO dataset (Lin et al., 2014) as the starting point for the model training to reduce training effort. The initial training process used 4,000 manual water tank labels and 20,000 training iterations. The corresponding training hyperparameters, also used for later model fine-tuning through semi-supervised self-training, were listed in Table 1.

Table 1
Hyperparameters used for training.

Batch Size	24
Learning Rate	0.0004
IoU Threshold	0.5
Optimizer	RMSprop
Optimizer Momentum	0.9
Optimizer Decay	0.9
Optimizer Epsilon	0.1

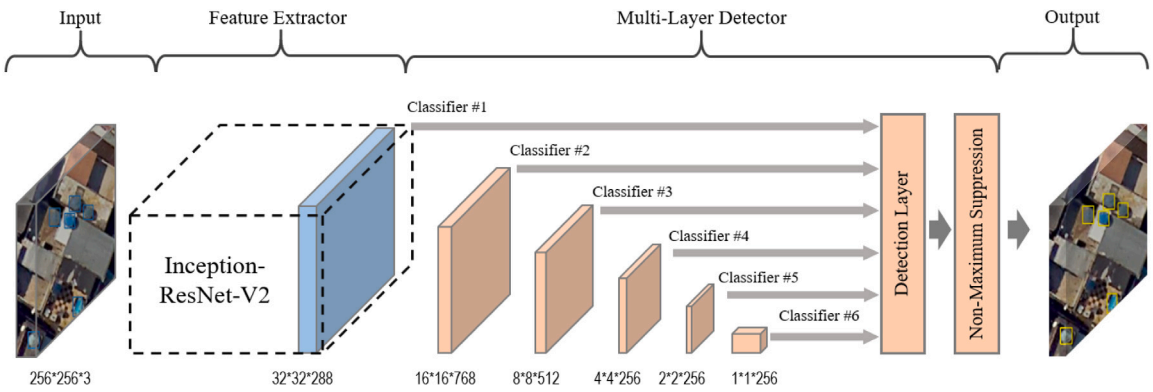


Fig. 4. Single-Stage Object Detection Network consisting of Inception-ResNet-V2 as feature extractor and multi layer detector with Non-Maximum Suppression layer as used for water tank detection models. The numbers at the bottom describes the dimension of the raster bands used at the different stages of the network. The output consists of bounding boxes for detected water tanks together with a confidence scale.

2.2.3. Pseudo water tank labeling

After the initial training process we generated additional pseudo water tank labels. Therefore, we applied the initially trained supervised model to predict water tanks in an unlabeled region. The selected region was approximately 15,000 ha and included all urban structure types, not just favelas and residential areas such as the manual water tank labeling process. Only water tank predictions with high confidence were used as additional pseudo labels to further fine-tune the water tank detection model. In our study, a confidence level of 80% was used as a lower threshold. This resulted in 10,800 additional water tank labels for the fine-tuning of our water tank detection model.

2.2.4. Semi-supervised self-training

The merged label set, combining the manual training labels and the pseudo labels, was used to fine-tune five new instances of our initial supervised model. The new instances differed in the number of semi-supervised training iterations in which the model was shown the additional pseudo-labels (cf. Fig. 5). For consistency, all new fine-tuned instances were trained for further 20,000 iterations. This resulted in five water tank detection models, all trained for 40,000 training iterations using the same number of manual water tank labels. All models were built upon the same initial supervised water tank detection model used for model inference and thus pseudo label creation. One of the new instances were fine-tuned without using the pseudo labels. This model was named the base model (BM). It was trained to evaluate the change in model performance reached through self-training.

2.2.5. Evaluation of semi-supervised self-training

Performance between all five water tank detection models was evaluated based on precision, recall, and their harmonic mean, the F1-score. Precision is defined as the ratio of the true positive objects to all detected objects. Recall describes the fraction of relevant objects that are successfully retrieved. The performance indicators were calculated based on the comparison between the intersection of the bounding boxes of the predictions and of the validation labels. The level of agreement of the two boxes was based on the Intersection over Union (IoU) value. The IoU takes values between 0 and 1. For a value of 0 the two boxes do not overlap at all. For a value of 1 they overlap completely. An IoU Value of 0.5 or higher for a detected object was considered a true positive. An IoU value lower than 0.5 as a false positive. In order to measure generalization capabilities of the models, we compared the change in model performance, in terms of the F1 score, between test sets from urban structures used in the manual label set and urban structures not used in the manual label set (see Fig. 6).

2.2.6. Large-scale predictions

The model with the best F1-score was used to predict water tanks for the whole metropolitan area of Rio de Janeiro. The prediction used over 10 million satellite image patches in parallel tasks. For data management we used the mapproxy API (Omniscale GmbH & Co. K.G., 2022). This allowed to store satellite imagery in subset folder structure. As for the detected water tanks, we pushed our predictions on each image patch to a PostGIS database. The database was then used for a post-processing step of filtering predictions by confidence scores.

2.2.7. Evaluation of water tank density as entomological proxy for Aedes Aegypti

To quantitatively evaluate our RQ2 of how well water tank density can capture the inner-urban distribution of *Aedes Aegypti* measured by entomological surveillance data, we ran a negative binomial generalized linear regression model (GLM) with a log-link function (Hilbe, 2012). The negative binomial GLM was selected as it allows the model to account for the overdispersion present in the count data. Corresponding equations are defined in Formula (1). As our response variable we selected the "mean eggs per trap" (MET) rate. As our explanatory variable we used water tank counts surrounding ovitrap locations in the range of 200 m which corresponds to the estimated *Aedes Aegypti* flight range. We used all 1,207 ovitrap locations with information on the MET rate.

$$\begin{aligned} MET_i &\sim NB(\hat{\mu}_i, \hat{\theta}) \\ \mathbb{E}(MET_i) &= \hat{\mu}_i * (1-\hat{\theta})/\hat{\theta} \\ \mathrm{Var}(MET_i) &= \hat{\mu}_i * (1-\hat{\theta})/\hat{\theta}^2 \\ \log(\hat{\mu}_i) &= \hat{\beta}_0 + \hat{\beta}_1 * Watertank_i \end{aligned} \tag{1}$$

3. Results and discussion

3.1. Comparison of training strategies

The results shown here are the outcome of the developed semi-supervised self-training approach for the large-scale detection of water tanks in the city of Rio de Janeiro (cf. Fig. 1). The workflow consisted of three major configuration points: First the targeted selection of suitable SSST regions, secondly the choice of an appropriate confidence threshold for pseudo label filtering and thirdly the testing of various ratios of supervised and semi-supervised training iterations (cf. Fig. 5). Sensitivity analysis for changing the first two configuration points were excluded from this result section. However, these parameters were chosen carefully during experimental design to enable model improvements through self-training. During experimental design, SSST regions for model inference were chosen to be small enough to save computational time, but large enough to generate a sufficient amount of pseudo labels required to fine-tune the object detection model.

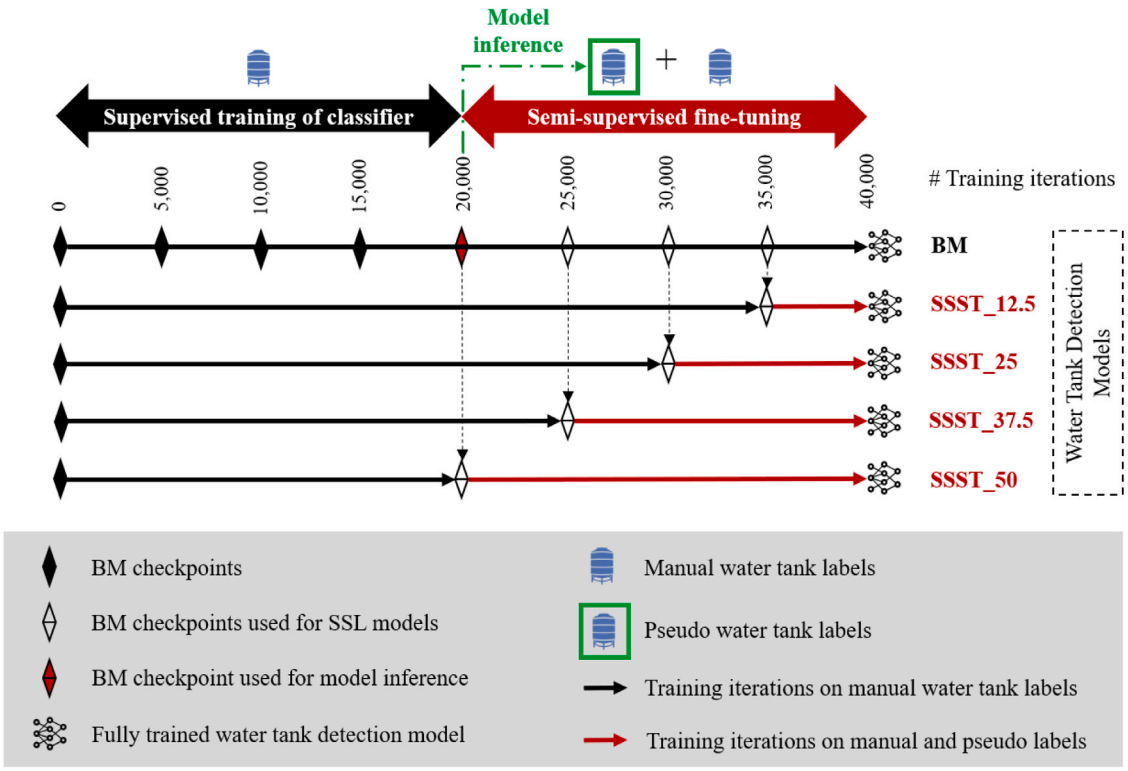


Fig. 5. Overview of trained models, water tank label sets used for training, and corresponding length of training epochs. BM = supervised base model, SSST — semi-supervised self-trained models.

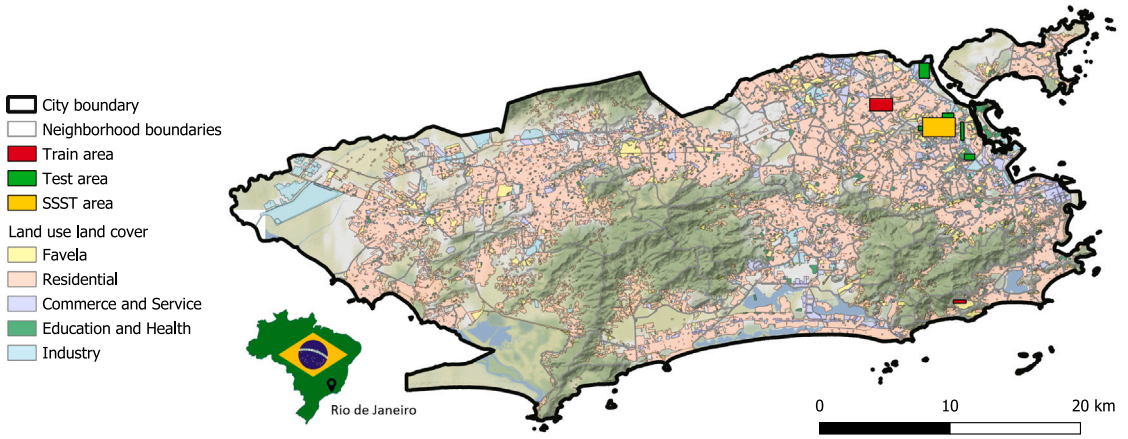


Fig. 6. The Land use land cover (LULC) in the city of Rio de Janeiro with train, test and semi-supervised self-training (SSST) region and topography in the background.

In addition, SSST regions were selected to cover all urban structure types present in the city of Rio de Janeiro to ensure a robust object detection for large-scale predictions. The confidence threshold for the filtering of model inference outcomes and thus for the generation of pseudo labels was chosen with respect to the model performance of our supervised base model with 20,000 training iterations. The results for the third configuration point of the workflow, namely the variation of different ratios of supervised to semi-supervised training iterations, were analyzed in more detail and discussed in the following.

An increasing training time on the merged label set of manual and pseudo water tank labels continuously improved the F1-score of our object detection models (Table 2). The best water tank detection model was the model that used the additional pseudo labels for the longest SSST time (50% of the total 40,000 iteration, SSST-50) with an overall F1-score of 0.84 averaged over all test labels. This significant improvement of 22% compared to the supervised base model indicates a good balance of precision and recall. All SSST models showed a slightly decrease in recall compared to the supervised base model — i.e the proportion of correctly detected water tanks to the sum of all true water tanks decreased. However, this was countered by major improvements in precision, as the amount of correct water tank predictions on all predictions was higher for all four SSST models used.

For the best (SSST-50) model, the relative increase in F1-score (cf. Table 3) was more obvious for urban structure types excluded in the manual label set (e.g. Commerce and Service, Education and Health, Industry) than for the urban structure types included in the manual

Table 2
Performance metrics of trained water tank detection models. The SSST-12.5/25/37.5/50 model used the pseudo labels during 12.5/25/37.5/50% of the 40,000 training iteration respectively.

Models	Precision(%)	Recall (%)	F1	% F1 Improvement
BM	**0.59**	**0.85**	**0.69**	–
SSST-12.5	0.84	0.69	0.76	+10%
SSST-25	0.9	0.7	0.79	+14%
SSST-37.5	0.86	0.78	0.82	+19%
SSST-50	0.86	0.82	0.84	+22%

label set (Favela, Residential). The F1-score improved, however, for all urban structure types. This makes the SSST-50 model more applicable for large-scale predictions than the supervised base model. These results were consistent with our expectations, namely that SSST models benefit from the additional knowledge collected by the machine itself, leading to more precise and robust water tank predictions across different urban structures relevant for large-scale predictions. The trained SSST-50 model is provided in the supplementary materials of this work.

However, we also identified several limitations in the results. First, not all urban structure types were used for model evaluation. We focused only on five of eleven land classes included in the LULC map where we expect human population and thus the largest risk for infections by *Aedes Aegypti*. The second limitation results from the manual labeling process. We generated the test set on the basis of satellite imagery instead of a field study. Non-visible water tanks underneath shelters were thus not included in our test labels for model evaluation. However, we assume that a field study for the labeling of water tanks would not mitigate the achieved performance improvement of the semi-supervised self-training approach. Much more likely, it would have an impact on the absolute performance metrics, but to the same extent on those of the supervised BM as on those of the SSST models. The third limitation of our study is the low amount of manual training labels (4,000) compared to the amount of pseudo labels used for training (10,800). This implies a relatively high risk of an inappropriate training with potentially incorrect pseudo labels which can accumulate the error in the iterative self-training process. To reduce such a bias, one could either develop a debiased self-training algorithm similar to the one proposed by Chen et al. (2022) or apply co-training of classifiers originally proposed by Blum and Mitchell (1998).

Further limitations of our study become apparent when visually inspecting raw prediction images of the SSST-50 model (cf. Fig. 7). Common false negative predictions included water tanks in the shade or partial shade. To minimize the amount of these false negative predictions one could further fine-tune the SSST-50 model by feeding it with more shaded water tank labels. It is noteworthy that the number of objects in our study area which appear similar to water tanks was quite high resulting in high numbers of false positives. While similar objects such as blue cars and rooftop ventilators were rarely labeled as water tanks by our models, circular water pools or blue sunshades on beaches were frequently false positives. The false positive detection of water pools could be solved by applying a size filter. The detection of blue sunshades on beaches could be eliminated by applying an automatic land use map based filtering. However, these solution methods would only work to a limited extent. For very small water pools and blue sunshades not located on beaches this solution method becomes obsolete. Another solution would be the filtering of predictions by confidence score as applied during SSST training.

Further improvements of our models might be achieved by changing parameters of our semi-supervised self-training framework. This includes the size of the areas used for supervised model predictions to generate pseudo labels, the confidence threshold score applied for pseudo label filtering, the overall training time for object detection models, and the corresponding point for conducting the semi-supervised self-training loop. The training of a two-stage object detector like Faster-RCNNs as proposed by Cao et al. (2019) could also be an option for further investigations.

An alternative method for the reduction of manual labeling effort for object detection could be data augmentation. Data augmentation describes the technique of increasing the training set by creating slightly modified copies of provided training samples, for example by changing the rotation of the label (Shorten and Khoshgoftaar, 2019; van Dyk and Meng, 2001). It is a widely used method especially applied to avoid overfitting. However, for our use case of generating a robust model for large-scale predictions over various urban structure types, semi-supervised self-training seems to be more suitable. Instead of creating label copies, self-training can create completely new water tank labels (pseudo label) that can appear in different shape, color, and with varying shadow coverage. In addition, it allows to incorporate background features in the training process, like different rooftop types or water tank densities, not necessarily present in the limited manual label set used. All these additionally features gathered during pseudo label generation via self-training are extremely relevant, when trying to train robust object detector using Convolutional-Neural-Networks (CNNs). Especially for applying these models on over 10 million satellite image patches covering all types of urban structures. Semi-supervised self-training can avoid overfitting similar to data augmentation (Nartey et al., 2020). Of course, do both methods, data augmentation and self-training, allow a cost-sensitive creation of additional labels, which is relevant for our use case to minimize the manual labeling effort and associated cost and time. However, the capability of learning additionally background features, not present in the manual label set, is only possible through self-training in an automatic manner. Nevertheless, self-training requires a relatively high configuration effort to be successful compared to data augmentation techniques as described in the beginning of this result section.

3.2. Modeling of urban mosquito abundance

As a highlight of this work, water tanks predicted by the SSST-50 model were distributed throughout the whole metropolitan area of Rio de Janeiro with a high spatial variability (cf. Fig. 8). The occurrence of water tanks was strongly dependent on inhabited areas. Forest areas were almost empty of water tanks. Water tank density within single neighborhoods also varied strongly. In addition to Fig. 8, we provide a raster layer with a spatial resolution of 200 m in the supplementary materials of this work. This consists of a raster value for water tank counts with confidence score above 90 percent and is intended to represent the spatial variance of water tanks at the resolution of an

Table 3
Goodness of fit indicators for the base model and the best performing SSST model for different urban structures. The performance was based on independent test data points.

Method	BM			SSST-50			
	Precision(%)	Recall(%)	F1	Precision(%)	Recall(%)	F1	% F1 Improvement
Favela	0.63	0.85	0.73	0.87	0.78	0.83	+14%
Residential	0.59	0.86	0.70	0.81	0.82	0.82	+17%
Industry	0.51	0.85	0.64	0.8	0.80	0.8	+25%
Education and Health	0.54	0.93	0.68	0.91	0.91	0.91	+34%
Commerce and Service	0.66	0.76	0.71	0.90	0.78	0.84	+18%
Average weighted by instance	0.59	0.85	0.69	0.86	0.82	0.84	+22%

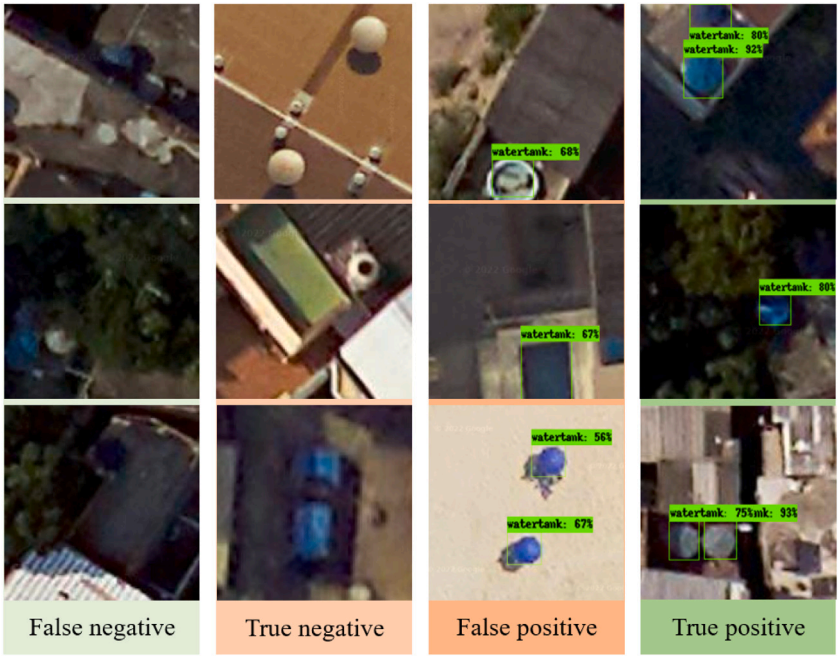


Fig. 7. Example for false negative, true negative, false positive, and true positive water tank predictions. Water tanks identified by the best performing SSST model are indicated by green bounding boxes together with the confidence of the prediction.

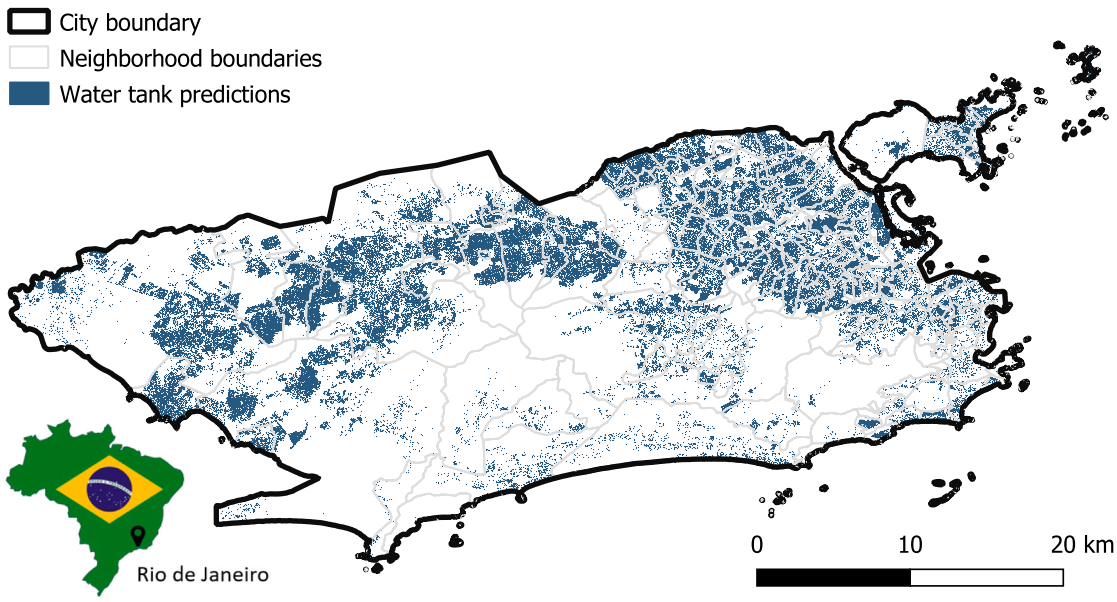


Fig. 8. Water tanks predicted by the best performing SSST model for the case study region. For orientation, the administrative boundaries of the neighborhoods are overlaid. The inset map to the lower left indicates the position of Rio de Janeiro in Brazil.

Table 4
Coefficients, standard errors, and p-values for negative binomial generalized linear regression model. Regression coefficients and standard errors are reported at the link scale. The dispersion parameter θ indicates underdispersion.

	Estimate	Std. Error	p-value
Intercept	1.535	0.709	$< 2e-16$
water tank count in 200 m ovitrap buffer	0.058	0.014	$< 2e-16$
θ	0.649	0.027	–

estimated *Aedes Aegypti* flight range for the whole city of Rio de Janeiro which could be used for urban mosquito modeling.

The results of our negative binomial generalized linear regression model (cf. Table 4) indicated that water tank density was a highly significant proxy for modeling the *Aedes Aegypti* MET rate. This was in line with our expectations and implicates that water tank density maps can be a useful indicator to enrich entomological surveillance data and thus support future vector control by providing more continuous and high resolution insights for urban mosquito distributions. The explained deviance for this regression model was 0.11. It was measured by Cohen's pseudo-R^2 (Cohen, 2013) (cf. Formula (2)) indicating that about 11% of the deviance in the response are explained by the model. The deviance function of the negative binomial GLM captured the increasing variance with the mean that is typical for count data.

The dispersion parameter captured thereby how strong the variance increases with the mean relative to a Poisson GLM, for which the variance equals the mean. The theta value of 0.649 corresponded to a significant overdispersion. This can be explained by the large number of zero values in the entomological dataset, which is why a negative binomial GLM was applied. Another reason for this is the low number of predictors used to model urban Aedes Aegypti distribution. However, other potentially relevant predictors have deliberately not been included in the model, which also explains the low value of the explained deviance. The addition of further explanatory variables is planned for follow-up activities.

$$Cohen's\ pseudo\ R^2 = 1 - \frac{model\ deviance}{null\ model\ deviance}$$
$$Negative\ binomial\ model\ deviance = 2\sum(y \cdot log(\frac{y}{\mu}) - (y + k^{-1})log(\frac{y+k^{-1}}{\mu+k^{-1}})) \quad (2)$$

4. Conclusion

The emergence of open-accessible big spatial data in combination with modern computing technologies has great potential to revolutionize the treatment of emergent infectious diseases transmitted by *Aedes Aegypti*. Especially for those that are missing effective vaccines and are therefore treated mainly by local vector control, namely dengue, Zika and chikungunya which cause thousands of deaths each year. In this paper we demonstrated how deep learning based object detection models in combination with open-accessible satellite imagery can be applied to extract a fine-grained and informative proxy for the urban modeling of *Aedes Aegypti* distribution, namely water tanks. Such models are essential to derive more targeted vector control interventions, allow cost savings in entomological surveillance, and most importantly a more efficient overall disease control. The results of this paper indicate that the burden of manual labeling necessary for large-scale and robust water tank detection can be substantially reduced by the development of a semi-supervised self-training workflow without compromising model accuracy. This increases not only real world applicability of our water tank detection model, but also its robust transferability to other *Aedes Aegypti* endemic cities where water tanks are common mosquito breeding sites. The measured significance in the association between water tank density and abundance of *Aedes Aegypti* showed the potential of the generated indicator to augment entomological surveillance gaps that occur when limited mosquito flight ranges are considered. The developed urban-specific indicator can thus bring novel insights into the high spatial variability of urban *Aedes Aegypti* distributions that can hardly be explained by commonly used low resolution features for *Aedes Aegypti* mapping. However, as the abundance of *Aedes Aegypti* depends on other predictors such as climatic conditions, upcoming research will explore the predictive power of water tank density in combination with these indicators. With these combined models for the fine-scale mapping of *Aedes Aegypti* distributions we hope to reveal hidden patterns not only with regard to urban *Aedes Aegypti* populations, but also for inner-urban pathogen transmission for dengue, Zika, and chikungunya. With these major contributions of our interdisciplinary research we hope to create new pathways for the science of computational eco-epidemiology and provide useful datasets as well as methods to public health authorities especially in the city of Rio de Janeiro, Brazil.

CRediT authorship contribution statement

Steffen Knoblauch: Conceptualization, Methodology, Software, validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Hao Li:** Conceptualization, Software, Writing – review & editing. **Sven Lautenbach:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Yara Elshiaty:** Software. **Antônio A. de A. Rocha:** Resources. **Bernd Resch:** Funding acquisition. **Dorian Arifi:** Funding acquisition. **Thomas Jänisch:** Funding acquisition. **Ivonne Morales:** Funding acquisition. **Alexander Zipf:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to https://doi.org/10.11588/data/7LLXFP. It contains our best water tank detection model, manual water tank labels used for training and testing as well as a high-resolution raster of water tank density for the city of Rio de Janeiro.

Acknowledgments

The authors would like to take this opportunity to thank the editors and reviewers for their valuable comments and suggestions. The entomological data was provided by the health ministry of the city of Rio de Janeiro. This work was funded by the Deutsche Forschungsgemeinschaft (DFG), Germany[grant number 451956976]. Sven Lautenbach acknowledges funding by the Klaus-Tschira Stiftung, Germany. The authors acknowledge support by the High Performance and Cloud Computing Group at the Center for Data Processing of the University of Tübingen, the state of Baden-Württemberg through bwHPC, and the German Research Foundation (DFG), Germany through grant no INST 37/935-1 FUGG.

Appendix A. Supplementary data

The supplementary data to this article can be found online at: https://doi.org/10.11588/data/7LLXFP. It contains our best water tank detection model, manual water tank labels used for training and testing as well as a high-resolution raster of water tank density for the city of Rio de Janeiro.

References

Agarwal, A., Chaudhuri, U., Chaudhuri, S., Seetharaman, G., 2014. Detection of potential mosquito breeding sites based on community sourced geotagged images. http://dx.doi.org/10.1117/12.2058121.

Amorim, J.H., Birbrair, A., 2022. Dengue vaccines: Where are we now and where we are going? Lancet Infect. Dis. 22 (6), 756–757. http://dx.doi.org/10.1016/S1473-3099(21)00753-2.

Andersson, V.O., Ferreira Birck, M.A., Araujo, R.M., 2018. Towards predicting dengue fever rates using convolutional neural networks and street-level images. In: 2018 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8. http://dx.doi.org/10.1109/IJCNN.2018.8489567.

Banerjee, S., Aditya, G., Saha, G.K., 2015. Household wastes as larval habitats of dengue vectors: Comparison between urban and rural Areas of Kolkata, India. PLoS One 10 (10), e0138082. http://dx.doi.org/10.1371/journal.pone.0138082.

Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. In: Bartlett, P., Mansour, Y. (Eds.), Proceedings of the Eleventh Annual Conference on Computational Learning Theory. COLT' 98, ACM Press, New York, New York, USA, pp. 92–100. http://dx.doi.org/10.1145/279943.279962.

Bomfim, R., Pei, S., Shaman, J., Yamana, T., Makse, H.A., Andrade, J.S., Lima Neto, A.S., Furtado, V., 2020. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. J. R. Soc., Interface 17 (171), 20200691. http://dx.doi.org/10.1098/rsif.2020.0691.

Boser, A., Sousa, D., Larsen, A., MacDonald, A., 2021. Micro-climate to macro-risk: Mapping fine scale differences in mosquito-borne disease risk using remote sensing. Environ. Res. Lett. 16 (12), 124014. http://dx.doi.org/10.1088/1748-9326/ac3589.

Bowman, L.R., Runge-Ranzinger, S., McCall, P.J., 2014. Assessing the relationship between vector indices and dengue transmission: A systematic review of the evidence. PLoS Negl. Trop. Dis. 8 (5), e2848. http://dx.doi.org/10.1371/journal.pntd.0002848.

Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., Liu, Y., Wu, Z., 2019. An improved faster R-CNN for small object detection. IEEE Access 7, 106838–106846. http://dx.doi.org/10.1109/ACCESS.2019.2932731.

Caputo, B., Manica, M., Filipponi, F., Blangiardo, M., Cobre, P., Delucchi, L., de Marco, C.M., Iesu, L., Morano, P., Petrella, V., Salvemini, M., Bianchi, C., Della Torre, A., 2020. ZanzaMapp: A scalable citizen science tool to monitor perception of mosquito abundance and nuisance in Italy and beyond. Int. J. Environ. Res. Public Health 17 (21), http://dx.doi.org/10.3390/ijerph17217872.

Chang, A.Y., Parrales, M.E., Jimenez, J., Sobieszczyk, M.E., Hammer, S.M., Copenhaver, D.J., Kulkarni, R.P., 2009. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. Int. J. Health Geogr. 8, 49. http://dx.doi.org/10.1186/1476-072X-8-49.

Chavasse, D., 2002. Know your enemy: Some facts about the natural history of Malawi's Anopheles mosquitoes and implications for malaria control 14 (1). pp. 7–8, URL https://pubmed.ncbi.nlm.nih.gov/27528915/.

Chen, B., Jiang, J., Wang, X., Wan, P., Wang, J., Long, M., 2022. Debiased self-training for semi-supervised learning. http://dx.doi.org/10.48550/arXiv.2202.07136, arXiv, URL https://arxiv.org/abs/2202.07136.

Cho, H., Low, R.D., Fischer, H.A., Storksdieck, M., 2021. The STEM enhancement in earth science "mosquito mappers" virtual internship: Outcomes of place-based engagement with citizen science. Front. Environ. Sci. 9, http://dx.doi.org/10.3389/fenvs.2021.682669.

Cohen, 2013. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Routledge, http://dx.doi.org/10.4324/9780203774441.

Colón-González, F.J., Sewe, M.O., Tompkins, A.M., Sjödin, H., Casallas, A., Rocklöv, J., Caminade, C., Lowe, R., 2021. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: A multi-model, multi-scenario intercomparison modelling study. Lancet Planet. Health 5 (7), e404–e414. http://dx.doi.org/10.1016/S2542-5196(21)00132-7.

Cunha, H.S., Sclauser, B.S., Wildemberg, P.F., Fernandes, E.A.M., Dos Santos, J.A., Lage, M.d.O., Lorenz, C., Barbosa, G.L., Quintanilha, J.A., Chiaravalloti-Neto, F., 2021. Water tank and swimming pool detection based on remote sensing and deep learning: Relationship with socioeconomic level and applications in dengue control. PLoS One 16 (12), e0258681. http://dx.doi.org/10.1371/journal.pone.0258681.

Da Queiroz, E.R.S., Medronho, R.d.A., 2022. Overlap between dengue, Zika and chikungunya hotspots in the city of Rio de Janeiro. PLoS One 17 (9), e0273980. http://dx.doi.org/10.1371/journal.pone.0273980.

Da Silva, Jr., J.B., Siqueira, Jr., J.B., Coelho, G.E., Vilarinhos, P.T., Pimenta, Jr., F.G., 2002. Dengue in Brazil: Current situation and prevention and control activities. URL https://iris.paho.org/handle/10665.2/32758.

Dandabathula, G., 2019. Automatic detection of overhead water tanks from satellite images using faster-RCNN. Int. J. Adv. Res. Comput. Sci. 10 (5), 34–37. http://dx.doi.org/10.26483/ijarcs.v10i5.6466.

David, M.R., Lourenço-de Oliveira, R., de Freitas, R.M., 2009. Container productivity, daily survival rates and dispersal of Aedes aegypti mosquitoes in a high income dengue epidemic neighbourhood of Rio de Janeiro: Presumed influence of differential urban structure on mosquito biology. Mem. Inst. Oswaldo Cruz. 104 (6), 927–932. http://dx.doi.org/10.1590/S0074-02762009000600019.

Dias, T.M., Alves, V.C., Alves, H.M., Pinheiro, L.F., Pontes, R., Araujo, G.M., Lima, A.A., Prego, T.M., 2018. Autonomous detection of mosquito-breeding habitats using an unmanned aerial vehicle. pp. 351–356. http://dx.doi.org/10.1109/LARS/SBR/WRE.2018.00070.

Ebi, K.L., Nealon, J., 2016. Dengue in a changing climate. Environ. Res. 151, 115–123. http://dx.doi.org/10.1016/j.envres.2016.07.026.

European Centre for Disease Prevention and Control, 2016. Aedes aegypti - Factsheet for experts. URL https://www.ecdc.europa.eu/en/disease-vectors/facts/mosquito-factsheets/aedes-aegypti.

Fernandes, E., Wildemberg, P., Dos Santos, J., 2020. Water tanks and swimming pools detection in satellite images: Exploiting shallow and deep-based strategies. pp. 117–122. http://dx.doi.org/10.5753/wvc.2020.13491.

Figueiredo, L.T.M., 2004. Dengue in Brazil: Past, Present and Future Perspective - Dengue Bulletin. Vol. 27. WHO Regional Office for South-East Asia, URL https://apps.who.int/iris/handle/10665/163881.

Franco dos Santos, S.A., Karam, H.A., Filho, A.J.P., Da Silva, J.C.B., Rojas, J.L.F., Suazo, J.M.A., Panduro, I.L.V., Peña, C.A.S., 2022. Dengue climate variability in Rio de Janeiro city with cross-wavelet transform. J. Environ. Prot. 13 (03), 261–276. http://dx.doi.org/10.4236/jep.2022.133016.

Getachew, D., Tekie, H., Gebre-Michael, T., Balkew, M., Mesfin, A., 2015. Breeding sites of Aedes aegypti: Potential dengue vectors in Dire Dawa, East Ethiopia. Interdiscipl. Perspect. Infect. Dis. 2015, 706276. http://dx.doi.org/10.1155/2015/706276.

Gibson, G., Souza-Santos, R., Pedro, A.S., Honório, N.A., Sá Carvalho, M., 2014. Occurrence of severe dengue in Rio de Janeiro: An ecological study. Rev. Da Soc. Brasileira de Med. Trop. 47 (6), 684–691. http://dx.doi.org/10.1590/0037-8682-0223-2014.

Google, 2022. TensorFlow object detection model zoo. URL https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md.

Gwitira, I., Murwira, A., Zengeya, F.M., Shekede, M.D., 2018. Application of GIS to predict malaria hotspots based on Anopheles arabiensis habitat suitability in Southern Africa. Int. J. Appl. Earth Obs. Geoinf. 64, 12–21. http://dx.doi.org/10.1016/j.jag.2017.08.009.

Haas-Stapleton, E.J., Barretto, M.C., Castillo, E.B., Clausnitzer, R.J., Ferdan, R.L., 2019. Assessing mosquito breeding sites and abundance using an unmanned aircraft. J. Am. Mosquito Control Assoc. 35 (3), 228–232. http://dx.doi.org/10.2987/19-6835.1.

Haddawy, P., Wettayakorn, P., Nonthaleerak, B., Su Yin, M., Wiratsudakul, A., Schöning, J., Laosiritaworn, Y., Balla, K., Euaungkanakul, S., Quengdaeng, P., Choknitipakin, K., Traivijitkhun, S., Erawan, B., Kraisang, T., 2019. Large scale detailed mapping of dengue vector breeding sites using street view images. PLoS Negl. Trop. Dis. 13 (7), e0007555. http://dx.doi.org/10.1371/journal.pntd.0007555.

Harrington, L.C., Scott, T.W., Lerdthusnee, K., Coleman, R.C., Costero, A., Clark, G.G., Jones, J.J., Kitthawee, S., Kittayapong, P., Sithiprasasna, R., Edman, J.D., 2005. Dispersal of the dengue vector Aedes aegypti within and between rural communities. Am. J. Trop. Med. Hygiene 72 (2), 209–220. http://dx.doi.org/10.4269/ajtmh.2005.72.209.

Hilbe, J.M., 2012. Negative Binomial Regression. Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511973420.

Hladish, T.J., Pearson, C.A.B., Toh, K.B., Rojas, D.P., Manrique-Saide, P., Vazquez-Prokopec, G.M., Halloran, M.E., Longini, I.M., 2020. Designing effective control of dengue with combined interventions. Proc. Natl. Acad. Sci. USA 117 (6), 3319–3325. http://dx.doi.org/10.1073/pnas.1903496117.

Honório, N.A., Da Silva, W.C., Leite, P.J., Gonçalves, J.M., Lounibos, L.P., Lourenço-de Oliveira, R., 2003. Dispersal of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. Mem. Inst. Oswaldo Cruz. 98 (2), 191–198. http://dx.doi.org/10.1590/s0074-02762003000200005.

Kantor, I.N., 2018. Dengue, zika, chikungunya y el desarrollo de vacunas. Medicina 78 (1), 23–28, URL https://pesquisa.bvsalud.org/portal/resource/fr/biblio-894542.

Knerer, G., Currie, C.S.M., Brailsford, S.C., 2020. The economic impact and cost-effectiveness of combined vector-control and dengue vaccination strategies in Thailand: Results from a dynamic transmission model. PLoS Negl. Trop. Dis. 14 (10), e0008805. http://dx.doi.org/10.1371/journal.pntd.0008805.

Limkittikul, K., Brett, J., L'Azou, M., 2014. Epidemiological trends of dengue disease in thailand (2000–2011): A systematic literature review. PLoS Negl. Trop. Dis. 8 (11), e3241. http://dx.doi.org/10.1371/journal.pntd.0003241.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2014. Microsoft COCO: Common objects in context. http://dx.doi.org/10.48550/arXiv.1405.0312.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2015. SSD: Single shot MultiBox detector. http://dx.doi.org/10.48550/arXiv.1512.02325.

Lobo, N.F., Achee, N.L., Greico, J., Collins, F.H., 2018. Modern vector control. Cold Spring Harbor Perspect. Med. 8 (1), http://dx.doi.org/10.1101/cshperspect.a025643.

Lorenz, C., Chiaravalloti-Neto, F., de Oliveira Lage, M., Quintanilha, J.A., Parra, M.C., Dibo, M.R., Fávaro, E.A., Guirado, M.M., Nogueira, M.L., 2020. Remote sensing for risk mapping of aedes aegypti infestations: Is this a practical task? Acta Tropica 205, 105398. http://dx.doi.org/10.1016/j.actatropica.2020.105398.

Louis, V.R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y., Dambach, P., 2014. Modeling tools for dengue risk mapping - A systematic review. Int. J. Health Geogr. 13, 50. http://dx.doi.org/10.1186/1476-072X-13-50.

Low, R.D., Nelson, P.V., Soeffing, C., Clark, A., 2021. Adopt a pixel 3 km: A multiscale data set linking remotely sensed land cover imagery with field based citizen science observation. Front. Clim. 3, http://dx.doi.org/10.3389/fclim.2021.658063.

Low, R.D., Schwerin, T.G., Boger, R.A., Soeffing, C., Nelson, P.V., Bartlett, D., Ingle, P., Kimura, M., Clark, A., 2022. Building international capacity for citizen scientist engagement in mosquito surveillance and mitigation: The GLOBE program's GLOBE observer mosquito habitat mapper. Insects 13 (7), http://dx.doi.org/10.3390/insects13070624.

Machault, V., Yébakima, A., Etienne, M., Vignolles, C., Palany, P., Tourre, Y., Guérécheau, M., Lacaux, J.-P., 2014. Mapping entomological dengue risk levels in martinique using high-resolution remote-sensing environmental data. ISPRS Int. J. Geo-Inf. 3 (4), 1352–1371. http://dx.doi.org/10.3390/ijgi3041352.

McFeeters, S., 2013. Using the normalized difference water index (NDWI) within a geographic information system to detect swimming pools for mosquito abatement: A practical approach. Remote Sens. 5 (7), 3544–3561. http://dx.doi.org/10.3390/rs5073544.

Mehra, M., Bagri, A., Jiang, X., Ortiz, J., 2016. Image analysis for identifying mosquito breeding grounds. In: 2016 IEEE International Conference on Sensing, Communication and Networking. SECON Workshops, IEEE, pp. 1–6. http://dx.doi.org/10.1109/SECONW.2016.7746808.

Messina, J.P., Brady, O.J., Golding, N., Kraemer, M.U.G., Wint, G.R.W., Ray, S.E., Pigott, D.M., Shearer, F.M., Johnson, K., Earl, L., Marczak, L.B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T.W., Reiner, R.C., Hay, S.I., 2019. The current and future global distribution and population at risk of dengue. Nat. Microbiol. 4 (9), 1508–1515. http://dx.doi.org/10.1038/s41564-019-0476-8.

Microsoft, 2022. Zoom levels and tile grid. URL https://learn.microsoft.com/en-us/azure/azure-maps/zoom-levels-and-tile-grid?tabs=csharp.

Morrison, A.C., Gray, K., Getis, A., Astete, H., Sihuincha, M., Focks, D., Watts, D., Stancil, J.D., Olson, J.G., Blair, P., Scott, T.W., 2004. Temporal and geographic patterns of Aedes aegypti (Diptera: Culicidae) production in Iquitos, Peru. J. Med. Entomol. 41 (6), 1123–1142. http://dx.doi.org/10.1603/0022-2585-41.6.1123.

Municipality of Rio de Janeiro, 2022. Land use data. URL https://www.data.rio/apps/PCRJ::uso-do-solo-1/about.

Muñoz, J.P., Boger, R., Dexter, S., Low, R., 2020. Mosquitoes and public health: Improving data validation of citizen science contributions using computer vision. pp. 469–493. http://dx.doi.org/10.1007/978-3-030-17347-0_23.

Nartey, O.T., Yang, G., Wu, J., Asare, S.K., 2020. Semi-supervised learning for fine-grained classification with self-training. IEEE Access 8, 2109–2121. http://dx.doi.org/10.1109/ACCESS.2019.2962258.

Omniscale GmbH & Co. K.G., 2022. MapProxy. URL https://mapproxy.org/.

Pan American Health Organization, 2019. Technical document for the implementation of interventions based on generic operational scenarios for Aedes aegypti control; 2019. URL https://iris.paho.org/handle/10665.2/51652.

Paploski, I.A.D., Rodrigues, M.S., Mugabe, V.A., Kikuti, M., Tavares, A.S., Reis, M.G., Kitron, U., Ribeiro, G.S., 2016. Storm drains as larval development and adult resting sites for Aedes aegypti and Aedes albopictus in Salvador, Brazil. Parasites Vectors 9 (1), 419. http://dx.doi.org/10.1186/s13071-016-1705-0.

Passos, W.L., Araujo, G.M., de Lima, A.A., Netto, S.L., Da Silva, E.A.B., 2020. Automatic detection of Aedes aegypti breeding grounds based on deep networks with spatio-temporal consistency. http://dx.doi.org/10.48550/arXiv.2007.14863.

QGIS Development Team, 2022. QGIS. URL https://www.qgis.org/en/site/.

Rocklöv, J., Dubrow, R., 2020. Climate change: An enduring challenge for vector-borne disease prevention and control. Nat. Immunol. 21 (5), 479–483. http://dx.doi.org/10.1038/s41590-020-0648-y.

Rosenberg, C., Hebert, M., Schneiderman, H., 2005. Semi-supervised self-training of object detection models. http://dx.doi.org/10.1184/R1/6560834.v1.

Runge-Ranzinger, S., McCall, P.J., Kroeger, A., Horstick, O., 2014. Dengue disease surveillance: An updated systematic literature review. Trop. Med. Int. Health : TM IH 19 (9), 1116–1160. http://dx.doi.org/10.1111/tmi.12333.

Sallam, M.F., Fizer, C., Pilant, A.N., Whung, P.-Y., 2017. Systematic review: Land cover, meteorological, and socioeconomic determinants of Aedes mosquito habitat for risk mapping. Int. J. Environ. Res. Public Health 14 (10), http://dx.doi.org/10.3390/ijerph14101230.

Schenkel, J., Taele, P., Goldberg, D., Horney, J., Hammond, T., 2020. Identifying potential mosquito breeding grounds: Assessing the efficiency of UAV technology in public health. Robotics 9 (4), 91. http://dx.doi.org/10.3390/robotics9040091.

Schrauf, S., Tschismarov, R., Tauber, E., Ramsauer, K., 2020. Current efforts in the development of vaccines for the prevention of zika and chikungunya virus infections. Front. Immunol. 11, 592. http://dx.doi.org/10.3389/fimmu.2020.00592.

Schwartz, L.M., Halloran, M.E., Durbin, A.P., Longini, I.M., 2015. The dengue vaccine pipeline: Implications for the future of dengue control. Vaccine 33 (29), 3293–3298. http://dx.doi.org/10.1016/j.vaccine.2015.05.010.

Secretaria de Vigilância em Saúde, 2013. Levantamento rápido de índices para Aedes aegypti (liraa) para vigilância entomológica do Aedes. URL https://bvsms.saude.gov.br/bvs/publicacoes/levantamento_rapido_indices_aedes_aegypti.pdf.

Semenza, J.C., Rocklöv, J., Ebi, K.L., 2022. Climate change and cascading risks from infectious disease. Infect. Dis. Therapy 11 (4), 1371–1390. http://dx.doi.org/10.1007/s40121-022-00647-3.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), http://dx.doi.org/10.1186/s40537-019-0197-0.

Su Yin, M., Bicout, D.J., Haddawy, P., Schöning, J., Laosiritaworn, Y., Sa-Angchai, P., 2021. Added-value of mosquito vector breeding sites from street view images in the risk mapping of dengue incidence in Thailand. PLoS Negl. Trop. Dis. 15 (3), e0009122. http://dx.doi.org/10.1371/journal.pntd.0009122.

Taborda, A., Chamorro, C., Quintero, J., Carrasquilla, G., Londoño, D., 2022. Cost-effectiveness of a dengue vector control intervention in Colombia. Am. J. Trop. Med. Hygiene http://dx.doi.org/10.4269/ajtmh.20-0669.

Trewin, B.J., Parry, H.R., Pagendam, D.E., Devine, G.J., Zalucki, M.P., Darbro, J.M., Jansen, C.C., Schellhorn, N.A., 2021. Simulating an invasion: Unsealed water storage (rainwater tanks) and urban block design facilitate the spread of the dengue fever mosquito, Aedes aegypti, in Brisbane, Australia. Biol. Invasions 23 (12), 3891–3906. http://dx.doi.org/10.1007/s10530-021-02619-z.

Uusitalo, R., Siljander, M., Culverwell, C.L., Mutai, N.C., Forbes, K.M., Vapalahti, O., Pellikka, P.K., 2019. Predictive mapping of mosquito distribution based on environmental and anthropogenic factors in Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinf. 76, 84–92. http://dx.doi.org/10.1016/j.jag.2018.11.004.

van Dyk, D.A., Meng, X.-L., 2001. The art of data augmentation. J. Comput. Graph. Statist. 10 (1), 1–50. http://dx.doi.org/10.1198/10618600152418584.

Vasconcelos, D., Yin, M.S., Wetjen, F., Herbst, A., Ziemer, T., Förster, A., Barkowsky, T., Nunes, N., Haddawy, P., 2021. Counting mosquitoes in the wild. In: Proceedings of the Conference on Information Technology for Social Good. ACM, New York, NY, USA, pp. 43–48. http://dx.doi.org/10.1145/3462203.3475914.

Vezzani, D., 2007. Review: Artificial container-breeding mosquitoes and cemeteries: A perfect match. Trop. Med. Int. Health : TM IH 12 (2), 299–313. http://dx.doi.org/10.1111/j.1365-3156.2006.01781.x.

WHO, 2012. Global Strategy for Dengue Prevention and Control, 2012-2020. World Health Organization, Geneva, Switzerland, URL https://apps.who.int/iris/handle/10665/75303.

WHO, 2017. Global vector control response 2017–2030. URL https://www.jstor.org/stable/resrep35629.

WHO, 2022. Vector-borne diseases. URL https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases.

Wilke, A.B.B., Vasquez, C., Carvajal, A., Medina, J., Chase, C., Cardenas, G., Mutebi, J.-P., Petrie, W.D., Beier, J.C., 2020. Proliferation of Aedes aegypti in urban environments mediated by the availability of key aquatic habitats. Sci. Rep. 10 (1), 12925. http://dx.doi.org/10.1038/s41598-020-69759-5.

Wilson, M.E., 2011. Chapter 4 megacities and emerging infections: Case study of Rio de Janeiro, Brazil. In: Khan, O.A., Pappas, G. (Eds.), Megacities & Global Health. American Public Health Association, http://dx.doi.org/10.2105/9780875530031ch04.

Wilson, A.L., Courtenay, O., Kelly-Hope, L.A., Scott, T.W., Takken, W., Torr, S.J., Lindsay, S.W., 2020. The importance of vector control for the control and elimination of vector-borne diseases. PLoS Negl. Trop. Dis. 14 (1), e0007831. http://dx.doi.org/10.1371/journal.pntd.0007831.

Youssefi, F., Javad Valadan Zoej, M., Ali Hanafi-Bojd, A., Borahani Dariane, A., Khaki, M., Safdarinezhad, A., 2022. Predicting the location of larval habitats of Anopheles mosquitoes using remote sensing and soil type data. Int. J. Appl. Earth Obs. Geoinf. 108, 102746. http://dx.doi.org/10.1016/j.jag.2022.102746.

Yukich, J.O., Lengeler, C., Tediosi, F., Brown, N., Mulligan, J.-A., Chavasse, D., Stevens, W., Justino, J., Conteh, L., Maharaj, R., Erskine, M., Mueller, D.H., Wiseman, V., Ghebremeskel, T., Zerom, M., Goodman, C., McGuire, D., Urrutia, J.M., Sakho, F., Hanson, K., Sharp, B., 2008. Costs and consequences of large-scale vector control for malaria. Malar. J. 7, 258. http://dx.doi.org/10.1186/1475-2875-7-258.